DOCUMENT RESUME

ED 218 330 TM 820 371

AUTHOR Dawson, Judith A.
TITLE Qualitative Research Findings: What Do We Do to Improve and Estimate Their Validity?
PUB DATE Mar 82
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Analysis; *Data Collection; Educational Research; Evaluation Methods; Measurement Techniques; Research Design; *Research Methodology; Standards; *Validity
IDENTIFIERS Qualitative Data; *Qualitative Research; *Validity Research

ABSTRACT

This paper is based on the premise that relatively little is known about how to improve validity in qualitative research and less is known about how to estimate validity in studies conducted by others. The purpose of the study was to describe the conceptualization of validity in qualitative inquiry to determine how it was used by the author of a research study, and to present a revised conceptualization of validity and suggestions for improving and estimating it. Validity is defined as the adequacy of a research report or other datum as a representation of some phenomenon for a particular purpose. Suggestions for improving validity are given for the stages of research design, data collection, data analysis, and reporting. Estimating validity requires making informed judgments about the adequacy of a research report for a particular purpose and includes considerations of correspondence between the representation and studied phenomenon; appropriateness of focus, scope, and intensity; and understandability. The assessment of validity in qualitative inquiry appears to be more judgmental, more relative, and less precise and certain than in quantitative inquiry. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218330

QUALITATIVE RESEARCH FINDINGS: WHAT DO WE DO TO IMPROVE AND ESTIMATE THEIR VALIDITY?

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.A. Dawson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Judith A. Dawson

Research for Better Schools, Inc.
444 North Third Street
Philadelphia, PA 19123

March 1982

Paper presented at the annual meeting of the American Educational Research Association, New York, March 1982.

TM 820 371

QUALITATIVE RESEARCH FINDINGS: WHAT DO WE DO TO IMPROVE AND ESTIMATE THEIR VALIDITY?

As educational researchers and evaluators increasingly adopt the qualitative approaches of anthropology and sociology, they modify the established methods. Today's researchers usually have less time to immerse themselves in a setting. They may visit it for only several days or weeks (Louis, Rosenblum, and Molitor, 1981; Stake and Easley, 1978), but even if studies span a year or more, researchers are unlikely to spend much more than half of their working hours at the site. Under these conditions, researchers need techniques for gathering data efficiently. They cannot afford to spend weeks simply gaining entry and establishing rapport or acquiring background information. Teams of researchers may work together on a study, creating a need for techniques of gathering data that are comparable and can be encoded (in written words or numbers) and communicated readily. Many such methods have been developed, as evidenced by recent publications on qualitative research methodology (e.g., Firestone and Dawson, 1982; Guba and Lincoln, 1981; Patton, 1980).

While researchers who write about qualitative methods often discuss validity, relatively little is known about how to improve it; less is known about how to estimate validity in studies conducted by others. Yet, validity must be dealt with more deliberately when researchers spend less time in the field; they are not as likely to correct mistaken impressions as researchers who have many opportunities over long periods of time to repeatedly observe phenomena or ask informants about them.

This is a revision of my previous work on validity in qualitative inquiry (Dawson, 1979, 1980). Since the initial paper, I have attempted to improve validity in a study I helped conduct, estimated it in studies conducted by others, and become aware of additional strategies used by other researchers to improve or estimate it. Here, I will briefly describe the initial conceptualization of validity in qualitative inquiry and how I used it. Then, I will present a revised conceptualization of validity and suggestions for improving and estimating it. Finally, I will reflect on that.

The Initial Conceptualization of Validity

Validity was defined as the adequacy of a research report or other datum as a representation of some phenomenon for a particular purpose. That definition meant that:

1. Data from qualitative inquiry--whether in the form of initial impressions, reports of informants, field notes, or reports to research audiences--are intended to represent real phenomena to others, phenomena such as classrooms, educational programs, individuals, events, meanings, and so forth.

2. Data cannot represent phenomena perfectly. Even a photograph is two-dimensional and often does not show an entire phenomenon in complete detail. Researchers cannot attain perfect representation; rather, they should strive for adequacy.[1]

3. Adequacy must have a referrent--adequacy for what? The adequacy of representations should be judged in relation to their purposes.

Three requirements for validity were described: (a) Correspondence with reality. Similar to the traditional motion of validity, this referred to agreement between research data and that which they purportedly represent. (b) Appropriateness of focus, scope, and intensity. This requirement

---

[1] "Adequacy" was not meant in its "barely sufficient" sense. Rather, it was used as a descriptor that might be applied to a continuum ranging from excellent to barely sufficient. At the lower end of the continuum, a representation would be barely valid.

acknowledged the limitations of representations--researchers cannot study every aspect of a phenomenon completely or communicate everything they learn to others--and the importance of considering purposes when identifying research foci and deciding with what scope and intensity to study and present them. (c) Understandability. The rationale behind this requirement was that validity, as adequacy for a particular purpose, is similar to utility: a representation cannot be used by audience members who do not understand it.

Strategies were suggested for increasing the likelihood that data collected during a study and reports written at its end would be valid. Methods for using the conceptualization and strategies to estimate the validity of representations written by others were also suggested.

Application of the Initial Conceptualization

The conceptualization was used for two years in a research study of the process of educational change, the Local School Improvement (LSI) study. A major purpose of the study was to generate hypotheses about that process. The primary research methods were non-participant observation and unstructured interviewing. The fieldwork spanned three school years and was conducted by a team of researchers whose members changed over the course of the study. I was involved during the last two years of fieldwork and the year of analysis and writing. Generally, researchers were assigned to two sites and spent an average of one day a week visiting them.

The conceptualization and strategies, for the most part, were useful. They were most useful for improving correspondence with reality--for increasing the likelihood that study data were valid. For that purpose, an awareness of threats to validity and of research conditions that were

likely to increase validity were especially valuable. Knowledge of the relationship of focus, scope, and intensity to validity became increasingly useful as the study progressed and the extent to which the data were sufficient became more important. Understandability was of less concern because the major audience was other researchers, but as the audience was expanded to educational practitioners this requirement, too, became more important.

Another application of the conceptualization was to estimate validity in three well-known ethnographies (Smith and Geoffrey, 1968; Spradley, 1970; Wolcott, 1973). Validity was estimated for various purposes, including those stated by the authors as well as other hypothetical purposes. Information used in the estimations included the books cited above, articles about study methodology written by the authors, and published reviews of the books. Three major conclusions emerged from that application. First, the conceptualization provided a useful framework for assessing validity in qualitative inquiry conducted by others. Second, information about study methodology was critical to that assessment. Without extensive descriptions of methodology, estimating validity is much more difficult. Third, despite the increased knowledge about validity that the application yielded, much uncertainty remained.

An Elaboration of the Initial Conceptualization

My conceptualization of validity in qualitative inquiry remains generally the same. The experiences of using the conceptualization to improve and estimate validity supported the notion of it as the adequacy of a

representation for a given purpose. The importance of focus, scope, and intensity and of understandability were also supported. However, their roles--as well as the role of purpose--in validity did vary. In this section, the relationship of each to validity and the ways in which they vary will be described.

Purpose is important as a referrent when validity is assessed. Research reports or other descriptions cannot represent phenomena completely and perfectly; rather, their adequacy as representations is judged relative to purpose. Studies often have multiple purposes--for example, the purposes stated in a grant or contract; perceived by project officers, researchers, and participants; or intended by target audiences and other users. Research reports are more adequate for some purposes than others; hence, they have multiple validities.

Reports of the study I helped conduct were more valid for some uses, and some audiences, than others. A major purpose of the study was to generate hypotheses about the process of educational change; reports of the study were generally adequate for research audiences to use for that purpose. Some people who were involved in the projects studied, however, viewed the reports more as descriptions of their own efforts. The reports were not adequate for that purpose; they focused on theoretical concerns rather than attempting to holistically describe the change efforts.

Appropriateness of focus, scope, and intensity refers to the extent to which the boundaries of a study or report are appropriate for a particular purpose. That is, given the necessary limitations of research in

representing phenomena, how appropriate were the foci and the breadth and depth with which they were portrayed? Researchers who approach an educational setting or other phenomenon without surveys or tests that define research boundaries in advance have a lot of latitude in deciding what to study--what setting elements, problems, issues, and so forth to pursue, what breadth and depth to pursue them with, and what perspectives to use. The choices that researchers make will be important to the adequacy of study findings for various purposes as the following examples illustrate.

Spradley's (1970) ethnography of urban tramps focused on some aspects of their culture to a much greater extent than others. The book included extensive descriptions of tramps' knowledge about different types of tramps, sleeping places, and encounters with law enforcement agencies, but little about their knowledge of obtaining money, food, and alcohol. The focus, scope, and intensity of the ethnography did not permit an overall description of tramps' culture; the book, though excellant in many ways, was not a valid representation of tramp culture in general but only of limited aspects of it.

Wolcott's (1973) ethnography of a principal focused on one principal. While the ethnography seems to be a valid portrayal of the work life of that principal, it is not an adequate representation of the principalship, which Wolcott sometimes wrote was the study's purpose. Its focus was too narrow for that purpose; it did not include many types of principals--urban, rural, female, inexperienced, and part-time, for example.

Focus, scope, and intensity need to be considered throughout a study. When it is designed, some of its parameters will be established, though

their specificity and flexibility will vary. Some studies, such as evaluation and policy research studies, are intended to address rather specific purposes; others are more exploratory and purposes are initially defined rather broadly. Focus, scope, and intensity will be more important during the research design stage of the former type of study than the latter. When purposes are defined narrowly at the beginning of a study, researchers should identify the types of information needed for those purposes. However, when they are defined broadly, more specific purposes can emerge as a study progresses and researchers decide to focus on particular areas or questions. Parlett and Hamilton (1976) and Stake (1981) discussed "progressive focusing;" Stake described how issues in a study of science education evolved during the 2½ year period between the release of the RFP and the completion of the study report.

During the data collection and analysis stages, the focus, scope, and intensity of a study affects the extent to which sufficient data will be available later for reporting to audiences who will use the data. Again, the extent to which intended uses are specified in advance varies among studies and the amount of influence of focus, scope, and intensity on data collection and analysis varies with it. At the reporting stage, when different reports may be written for different purposes, the topics covered in each and their breadth and depth will influence the reports' validity for various purposes.

Understandability is important primarily at the reporting stage of a study. The validity or utility of a report for a given purpose or set of purposes is partially dependent on the extent to which the audience can

understand it. If they cannot, it is not valid for them. This means not only that audiences are able to comprehend a report's contents, that the technicality of the report and the level at which it is written are appropriate, but that the report is written so as to convey the type of understanding needed. If an audience needed to understand a described setting vividly, the report would contain a lot of pure, detailed description and quotes from participants. On the other hand, if the audience needed only general knowledge of the setting, the report could be more broad and abstract.

Improving Validity

Many strategies for improving validity are available. Some will be described here. They contribute to validity in different ways and have different limitations. Therefore, researchers need to use a wide variety of strategies. Suggestions for improving validity are arranged here by the research stage at which they will be used primarily: research design, data collection, data analysis, or reporting.

Research Design Stage

Many of the decisions that are made when qualitative studies are designed can contribute to the validity of the data that are collected. Such design decisions will, of course, be subject to change--as, for example, when researchers use new information to identify additional questions that should be explored--but validity should be taken into consideration during modifications as well as initial decisions.

Use an interacting team of researchers. Research teams, whether working in single or multiple sites, can often develop more holistic, less biased perceptions of phenomena than can researchers working alone. Team members can help clarify one another's interpretations, identify over-emphases and

other biases, and contribute different perspectives. Researchers in the LSI study had varying academic (sociology of education, anthropology, educational research and evaluation) and teaching (none, kindergarten, secondary school English) backgrounds, giving us different perspectives which all contributed to the final analysis. Of course, there must be provisions for interaction among team members. They can work in adjacent physical spaces when not in the field, as we did in the LSI study; hold retreats at a central location periodically; or circulate research memoranda. Researchers who are assigned to different sites should visit one another's sites.

Triangulate. Different research methods, informants, and contexts contain different sources of error. Using combinations of them can help reduce error. Specifying multiple data collection methods (observation, interviewing, open-ended questionnaires), types of informants (administrators, teachers, students), and situations (classrooms, faculty lounges, staff meetings) can help maximize triangulation (Denzin, 1970; Smith and Pohland, 1974).

Establish research boundaries. Tentatively identifying research boundaries can help ensure that appropriate data will be available at the end of the study. The boundaries may be changed later, but it may be useful to make initial decisions regarding research foci (setting aspects, questions, issues, etc.), the extent to which they will be examined in breadth and depth, and the persepctives that will be used to investigate them.

Develop a system for maintaining field notes. Researchers will need to review previous field notes--to refresh their memories of events, confirm the sequence in which they occurred, verify that an informant's current remarks do or do not contradict previous ones, and so on. Members of a team may want to consult one another's field notes. Provisons should be made for recording field notes--perhaps facilitating that with tape recorders for researchers to dictate into and secretaries to transcribe the notes--and making them accessible through coding or storage systems.

Employ external perspectives. In addition to using teams of researchers and triangulation, it is often useful to bring in outsiders with perspectives that are unlikely to be tapped otherwise. Subject matter or process specialists and experienced practitioners may view a setting differently than researchers or participants. Stake and Easley (1978), in a multi-site case study, used site-review teams that included experts to visit settings while or immediately after individual researchers conducted case studies there.

Allow sufficient time in settings. There are several advantages to studying a setting over a lengthy period of time. Researchers can collect more data to inform their interpretations (Adams, 1981: Greene and David, 1981), test conclusions many times in many ways (Becker, 1970), understand participants well enough to accurately interpret their meaning (Bruyn, 1966), and avoid conducting too many observations at unrepresentative times (Bogdan and Taylor, 1975).

Make follow-up site visits. Researchers sometimes consider it helpful to return to settings after beginning data analysis to collect additional data or refine or confirm the analysis. Yin and Gwaltney (1981) designed a multi-site study with two rounds of site visits so that they could use the second round to refine and confirm their analysis of data collected during the first round.

Establish confidentiality procedures. Interviewees sometimes withhold or distort information because they fear it may be used against them or their organization, as in negative evaluative statements about individuals or organizations or when interviewee's remarks are communicated to others, such as their superiors. Depending on the study, it is usually possible to develop guidelines that will help reduce such threats to validity. Study participants can be assured at the outset, for example, that individuals will not be evaluated and that their remarks will not go beyond the research team unless anonymously.

Data Collection Stage

Many factors can threaten validity in qualitative inquiry during the data collection stage. The threats that are described below are categorized by the point in the data collection process which they are most likely to influence.

> Exposure to, or opportunity to observe, phenomena. Researchers and informants sometimes have only partial access to settings--e.g., they may not have access to contacts with clients or interactions among administrators. That exposure may be brief and incomplete or unrepresentative. Spatial location may cause misinterpretation. An informant may rely on second- or third-hand reports rather than direct observation (Lofland, 1971).

Perception of phenomena and access to those perceptions. People cannot deal with all the information to which they are exposed; they tend, instead, to select that which is familiar or of interest and to screen out other information (Sadler, 1981; Trankel, 1972). Fieldworkers view phenomena from different perspectives and attend to some aspects of the phenomena more than others; informants do the same. People sometimes place undue emphasis on first impressions, have difficulty dealing with conflicting or missing information (Sadler, 1981), or have limited introspective access to their thoughts and behaviors--e.g., are unaware of whether or how a particular stimulus influenced a response (Reichardt, 1981).

Memory of observations or perceptions. Informants may have poor recollections, especially when they have suppressed or altered their memories to meet an emotional need or when a long period of time has elapsed (Dean and Whyte, 1969; McCall, 1969). Researchers' memories are especially likely to be faulty during the beginning phases of fieldwork when they are overwhelmed with new information or when they do not record it soon enough.

Interpretation of observations. Researchers and informants may be unknowingly influenced by their biases toward or against that which they observe. Researchers may over-identify with participants who share their biases and over-report the views of those participants while neglecting others (Becker, 1974; Wolcott, 1977). An observer's understanding of the context of an occurrence will influence how he or she interprets it. Some setting members have more savvy than others and more access to information and can place it in context and interpret it more usefully.

Communication of information to others. Informants sometimes have poor communication skills; they simply cannot effectively transfer what they know to others. Cultural differences may lead to miscommunication. Informant responses may be influenced by efforts to impress or maintain a social relationship with a researcher. Informants may deliberately mislead a researcher in order to protect themselves or their organizations (Van Maanen, 1979; Vidich and Bensman, 1960). During interviews, respondents may feel inhibited or constrained (Dean and Whyte, 1969; McCall, 1969).

To reduce the influence of those and other threats to validity, researchers can (a) be continuously alert to sources of invalidity, (b) carefully establish and monitor field relationships, (c) use confirmation strategies, and (d) carry out the provisions of the research design, modifying them as needed. Another strategy researchers can use during the data collection

stage to improve validity is to progressively define the research boundaries.

Be continuously alert to sources of invalidity. Perhaps one of the most effective things researchers can do to maximize validity is to remain constantly aware of threats to it and attempt to reduce or at least assess their effects. Researchers should know the types of threats that have been encountered by others and be alert to new ones. They should then attempt to reduce or counteract the threats; for example, researchers can seek more complete access to settings, select informants who seem especially knowledgeable or trustworthy, and suspend belief until findings are verified.

Carefully establish and monitor field relationships. Friendly, trusting relationships between researchers and study participants affect validity by influencing a researcher's access to information and the completeness and accuracy of informants' reports. While strategies for establishing good field relationships are somewhat setting- and person-specific, researchers should generally enter a setting gradually, eventually involving more setting members and seeking answers to more research questions; be open and honest about research purposes and methods; and assure informants of confidentiality procedures (Johnson, 1975; Schatzman and Strauss, 1973). Because field relationships change over time, they should be examined frequently.

Use confirmation strategies. Confirming, or attempting to disconfirm, findings is a very direct validation approach. Many different confirmation strategies exist.

Informally discuss observations and interpretations with setting members, research team members, and others who are familiar with that or similar settings. Frequent discourse of this sort can contribute to the progressive development of a coherent understanding of a setting as well as confirm or disconfirm portions of that understanding. Sometimes setting members become key informants whom researchers use regularly to test their impressions (Whyte, 1955) or research partners who make field notes, react to the senior researcher's notes, and work with the researcher to develop an interpretation of the setting (Smith and Geoffrey, 1968).

Keep notes of important findings and consider them tentative until they have been confirmed two or three times (Sutherland, 1975).

As data are jointly collected, coded, and analyzed, decide what kinds of data should be collected next to further inform the analysis and what group or subgroup is most likely to provide the data; repeat the process. Continue repeating it until the analysis is so complete that new data are no longer informative (Glaser and Strauss, 1967).

Hypothesize about conditions which would exist if findings were true and look for those conditions (Miles, 1979). Or, look for conditions that would not exist if the findings were true (Yin, 1981).

Systematically gather bits of evidence that will help inform and validate a conclusion. The evidence should fit together in a whole that constitutes "structural corroboration" of the conclusion (Eisner, 1979).

Implement the provisions of the research design. Many of the suggestions in the previous section on research design were that it include provisions for researchers to carry out various strategies for improving validity. Many of those strategies would be used in the data collection stage. In particular, researchers should interact with other team members, triangulate, keep field notes systematically, make arrangements for site visits by external consultants who would provide additional perspectives, spend sufficient time in settings, make follow-up site visits, and establish confidentiality procedures. Naturally, a researcher may need to modify the initial plans, and should do so.

Progressively define research boundaries. As issues emerge and researchers define their questions and purposes more narrowly, the focus, scope, and intensity needed to address them will become progressively more clear. Investigators can then decide where the data seem to be sufficient and where more data need to be collected--e.g., on what problems or issues, for what individuals, groups or organizational units.

My experience conducting qualitative inquiry indicates that many of these strategies can become routine behavior that researchers practice almost unconsciously during the data collection stage. I continually questioned the motives and reliability of informants, estimated the influence of their (and colleagues') special interests and biases on their interpretations, and identified people whose reports I had to verify independently before adding them to my stock of knowledge about a site. Colleagues and I offered rival interpretations of one another's observations. We discussed our perceptions with each other and with other people who knew the situations we were studying. We kept mental notes of data we suspected were invalid or incomplete; we corrected them during subsequent site visits. These behaviors may have contributed more to the validity of our data than anything else we did.

Data Analysis Stage

In qualitative inquiry, data collection and analysis are frequently not distinct. Researchers tend to interpret data while collecting them and iteratively develop overall explanations of phenomena. Although the most effective validation efforts may be those that occur before and during

data collection, while researchers can still return to a setting to gather more information, there are some strategies that researchers can use to improve validity after data collection has ended.

Check the frequency and distribution of phenomena. Systematically analyze field notes for multiple reports of a particular finding. For each instance, record type of data and circumstances: Statement by interviewee or observation by researcher? Statement made in group setting or to researcher alone? Statement volunteered or made in response to researcher's question? In general, according to Becker (1969), the existence of a phenomenon is supported more under certain conditions--when instances of it fall into multiple categories, volunteered statements outnumber directed statements, volunteered statements are made to the group as well as the researcher alone, and there is a reasonable proportion of activities to statements.

Reporting Stage

During the reporting stage, researchers can engage in three kinds of activities that contribute to validity. First, researchers can make certain that the contents of a report--its focus, scope, and intensity--are appropriate for its purposes and that it will be understandable to audiences. Second, they can submit drafts of the report to others for their assessments of validity. Third, researchers can include in the report information that will allow readers to estimate validity.

Make certain that report contents are appropriate and understandable. The foci included in a report and the depth and breadth with which they are covered should be appropriate for its purposes. Researchers need to decide

what setting aspects, questions, issues, and so forth to include as well as the level of detail, in-depth examination, and thoroughness of coverage. And, they need to decide how much contextual information to include.

Reports are likely to be more understandable if they are written clearly in a language the audience can comprehend and are capable of conveying the type of knowledge needed. For some audiences, difficult words, technical language, and educational jargon should be avoided. If an audience needs a vivid understanding of the described phenomena, the report should contain such things as pure, detailed description as well as quotes from setting members and from field notes. The understandability of reports can be assessed by asking audience members to read report drafts and later discussing their perceptions of the described setting.

Submit report drafts to others for their assessments of validity. Many researchers and evaluators ask selected setting participants to read drafts of reports and react to them (e.g., Stake and Easley, 1978). Reports can then be revised or reactions can be included in footnotes (Miles, 1978) or appendices (Alkin, Daillak, and White, 1979). Report reviewers might also include other people (colleagues, specialists, administrators, and so on) who are familiar with that or similar settings.

Include information that will allow readers to estimate validity. It is not sufficient that researchers consider their reports valid. Readers must also be able to estimate validity. They should not be asked to simply accept a report on faith; they should be given an opportunity to scrutinize a report and decide whether it is sufficiently valid for them. Also, they may want to use a report for other than the intended purposes and need to

judge validity for the new purposes. They may want to apply specific findings to new settings and to know about the validity of those findings. Researchers can do several things to help readers estimate validity.

> Insert data throughout the report to help readers appraise the researcher's interpretations. Verbatim quotes from respondents, detailed descriptions of occurrences, background demographic information (e.g., size of school, educational and experiential backgrounds of teachers), and quotes from field notes can be useful for that purpose. So can frequency distributions of the inclusion of a particular finding in the field notes. Data display charts that list events in sequence, summarize participant reactions, and so forth (Miles, 1980) not only help readers estimate validity but also make comprehension of the report easier for them and, during preparation, alert researchers to data that are missing or were not considered during interpretation.
>
> Write a report that is coherent, that conveys a story in which the parts corroborate one another and fit together to form a whole.
>
> Describe the methods used to collect data and to help ensure validity. Extensive methodological statements are often very important to readers who want to assess validity. They help convince the reader that the research was rigorous, that many precautions were taken. Reports that do not include descriptions of study methodology sometimes lack credibility to readers.
>
> Include written reflections on validity. Researchers' assessments of validity--where and for what purposes it seems to be very strong, where it is inconclusive, what kinds of evidence support or raise questions about validity, what threats to validity were present or absent and the effects of those that were present--can help the reader judge validity and convince him or her that the researcher attempted to maximize it. Huberman (1981) discussed the validity of the data for each of three sub-units included in his study. He noted, for example, that conflict existed at one site and participants tended to be evasive and to disagree with one another about occurrences; he gave evidence of validity in another site (e.g., interviews elicited redundant information, two informants critiqued preliminary write-ups) and identified where the data were thin.

The Need to Use Multiple Validation Strategies

The strategies for improving validity that have been presented here contribute to validity differently and have different weaknesses. For

example, confirmation is very direct strategy that appears to provide unmixed evidence of validity. However researchers by themselves cannot always confirm that their perceptions and interpretations were correct; they can repeatedly misunderstand something without realizing it. Because of that, they also cannot confirm an overall report. Participants can confirm a researcher's perceptions and interpretations as well as an overall report, but those confirmations may have several limitations. Participants do not always spend enough time studying a description to react to it comprehensively; they may be defensive and overreact or avoid mentioning information researchers overlooked; they may be too involved in a situation to view it as clearly as an outsider; and, they may disagree among themselves about validity.

Triangulation has become almost an expectation of qualitative inquiry. Different kinds of information can be collected most effectively using different methods; a combination of methods can provide a more comprehensive understanding of a setting. However, triangulation can elicit information that is difficult to reconcile and interpret. Data from different methods or sources can appear to contradict each other. Yet, all of the data may be valid; they may simply represent different views of phenomena or ways of looking at them.

Some strategies--such as establishing confidentiality procedures, being alert to sources of invalidity, and monitoring field relationships--increase the likelihood that data will be valid but do not provide other evidence that the data actually were valid. And, researchers cannot identify or eliminate all sources of invalidity, do not know specifically what

research conditions will be most favorable for validity in a given setting, and do not have perfect knowledge of or control over conditions in that setting.

Thus, researchers must use multiple strategies for improving validity. Preferably, the strategies will be combined carefully in an effort to optimize validity by capitalizing on their strengths and limitations. Research designs can promote validity by including provisions for teams of researchers, triangulation, early consideration of boundaries, and extensive time in settings, for example. Researchers can be continually aware of the importance of validity and of factors that might reduce it; deliberately attempt to confirm, and disconfirm, findings; use rigorous procedures to analyze data; and ask participants to react to report drafts. Researchers can also attempt to make reports as understandable as possible to intended audiences.

Estimating Validity

Estimating validity involves making informed judgments about the extent to which a research report or other type of representation appears to be adequate for a particular purpose. Such judgments should include consideration of correspondence between the representation and studied phenomenon; appropriateness of focus, scope, and intensity; and understandability. Some questions that might be asked when estimating validity include:

> Was the researcher aware of the need to deliberately take actions to improve validity? What did he or she do about it? What was the researcher's assessment of validity?

What was the researcher's background? How much prior research experience did he or she have? Did the researcher have experience in similar settings? What were his or her theoretical orientations? Biases? Pre-conceived impresessions of the setting?

How might the research conditions have influenced validity? What was the researcher's relationship with participants? Was the study threatening to them? Were confidentiality procedures followed? How much time did the researcher spend in the setting?

What confirmation strategies were used? Did the researcher return to the setting to deliberately seek confirmation or disconfirmation? Keep a record of findings that should be considered tentative until verified? Discuss his or her impressions with setting members while the research was in progress? Did setting members read and react to report drafts? Did the researcher subsequently revise the report or include the reactions in it?

Did the researcher use multiple data collection methods? Did he or she formally interview participants and talk with them informally and observe them interacting with colleagues? Did the researcher include a variety of participants who were likely to have different perspectives? How were data from various sources compared and reconciled?

Was more than one researcher familiar with the setting and involved in writing the report? Did the researchers interact with one another frequently?

How were data stored and retrieved? Were extensive field notes recorded? Was an indexing or categorizing system used to make them readily accessible during data analysis?

What consistencies and inconsistencies exist within the report? Do the data support the interpretations? Do some data appear to contradict other data?

Are the foci of the report--the setting aspects, questions, issues, and so forth--appropriate for a particular purpose? Are they described in sufficient breadth? Detail? Does the report provide the information needed?

Is the report understandable to the people who will use it? Is it written at an appropriate level? If for a lay audience, does it not contain too much educational jargon or technical language? If the audience needs to understand the setting vividly, does the report contain detailed descriptions and quotes from participants?

The major source of information used to answer the questions would probably be the research document itself, particularly its description of study methodology. Articles by the author, especially articles about the methodology, might also be useful.

Other kinds of information may also be available to consider in estimates of validity. Journal reviews of the study or report may include judgments of validity. The person who estimates validity may have personal knowledge of the setting or others similar to it, or access to other people with such knowledge. Reports of other studies of the setting or articles about it may include information to use in estimating validity.

Estimates of validity should be made on the basis of as much evidence as possible. It is especially important that information about study methodology is available, although estimates may be strengthened if other sources of information that are independent of the researcher are also available. The evidence should show that the researcher used multiple strategies to improve validity. For example, the researcher should have deliberately reduced the effects of threats to validity, established and maintained research conditions that were favorable to validity, used triangulation, and sought confirmation with participants. Perhaps most importantly, there should be evidence that the researcher was aware of the need to deliberately and repeatedly attempt to improve and monitor validity.

Some Reflections on Validity

The extent to which reports of qualitative inquiry are valid for given purposes can be improved and estimated, though with varying amounts of precision and certainty. Assessments of validity--whether by researchers

or research users--are in the form of descriptive statements, or unverbalized thoughts. They do not have the precision implied by numbers. They are likely, however, to allow one to distinguish between different amounts of validity, to know that one set of information has more or less validity for a particular purpose than another, and something about the magnitude of the differences.

Sometimes validity can be assessed with considerable confidence--if, for example, multiple validation strategies produced dependable evidence that consistently supported or denied the existence of validity. Often, however, judgments of validity are complex and made with uncertainty. The amount of evidence that is available may be insufficient. It may be inconsistent in its support or denial of validity. The evidence may suggest that parts of a report are quite valid but other parts are of low or indeterminate validity.

The assessment of validity in qualitative inquiry may appear to be more judgmental, more relative, and less precise and certain that in quantitative inquiry. Numerous systematic procedures for estimating validity in quantitative inquiry have been developed; many produce numerical estimates. However, quantitative researchers recommend that validity should be estimated using a variety of evidence that should be combined judgmentally (APA, 1974). Some of the evidence would probably be in the form of numbers and some would not; estimates of content validity, for example, are not assigned numbers. The evidence might contain some of the uncertainties discussed above. Also, quantitative researchers have recommended

that estimates of validity should be relative to intended use, as in qualitative inquiry (Cronbach, 1971). Saying that test developers should help users judge the validity of test interpretations, Cronbach (1980, p. 107) noted that uncertainties will remain.

> The validity of an interpretation cannot be established by a research monograph or detailed manual. The aim for the report is to advance sensible discussion. Why should we wish for more? On matters before the public, the evidence usually is clouded. The institutions of the polity are geared to weigh up reasonable, partly persuasive, disputed arguments; and they can be tolerant when we acknowledge uncertainties. The more we learn, and the franker we are with ourselves and our clientele, the more valid the use of tests will become. (emphasis in original)

Bibliography

Adams, K. A. The keen edged feather: Intuitive analysis and reporting of qualitative evaluation data. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

Alkin, M. C., Daillak, R., & White P. Using evaluations: Does evaluation make a difference? Beverly Hills, Calif.: Sage Publications, 1979.

American Psychological Association. Standards for educational and psychological tests. Washington, D.C.: author, 1974.

Becker, H. S. Problems of inference and proof in participant observation. In G. J. McCall & J. L. Simmons (Eds.), Issues in participant observation: A text and reader. Reading, Mass.: Addison-Wesley, 1969.

Becker H. S. Sociological work: Method and substance. Chicago: Aldine, 1970.

Becker, H. S. Whose side are we on? In G. Riley (Ed.), Values, objectivity and the social sciences. Reading, Mass.: Addison-Wesley, 1974.

Bogdan, R., & Taylor, S. J. Introduction to qualitative research methods: A phenomenological approach to the social sciences. New York: John Wiley & Sons, 1975.

Bruyn, S. T. The human perspective in sociology: The methodology of participant observation. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1971.

Cronbach, L. J. Validity on parole: How can we go straight? New Directions for Testing and Measurement, 1980, 5, 99-108.

Dawson, J. A. Validity in qualitative inquiry. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1979.

Dawson, J. A. Validity in qualitative inquiry. Unpublished doctoral dissertation, University of Illinois, 1980.

Dean J. P., & Whyte, W. F. How do you know if the informant is telling the truth? In G. J. McCall & J. L. Simmons (Eds.), Issues in participant observation: A text and reader.

Denzin, N. K. The research act: A theoretical introduction to sociological methods. Chicago: Aldine, 1970.

Eisner, E. W. The educational imagination. New York: Macmillan, 1979.

Firestone, W. A., & Dawson, J. A. Approaches to qualitative data analysis: Intuitive, procedural, and intersubjective. Philadelphia, Pa.: Research for Better Schools, Inc., 1981.

Glaser, B. G., & Strauss, A. L. The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine, 1967.

Greene, D., & David, J. L. A research design for generalizing from multiple case studies. Palo Alto: Bay Area Research Group, 1981.

Guba, E. G., & Lincoln, Y. S. Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic appraoches. San Francisco: Jossey-Bass Publishers, 1981.

Huberman, A. M. School-university collaboration supporting school improvement: The Midwestern State Teacher Center Network case (Vol. I). Washinton, D. C.: Knowledge Transfer Institute, Center for Technology and Administration, The American University, June 1981.

Johnson, J. M. Doing field research. New York: The Free Press, 1975.

Lofland, J. Analyzing social settings: A guide to qualitative observation and analysis. Belmont, Calif.: Wadsworth, 1971.

Louis, K. S., Rosenblum, S., & Molitor, J. A. Linking R&D with schools: Strategies for knowledge use and school improvement. Washington, D. C.: National Institute of Education, July 1981.

McCall, G. J. Data quality control in participant observation. In G. J. McCall & J. L. Simmons (Eds.), Issues in participant observation: A text and reader. Reading, Mass.: Addison-Wesley, 1969.

Miles, M. B. Qualitative data as an attractive nuisance: The problem of analysis. Administrative Science Quarterly, 1979, 24, 590-601.

Miles, M. B. Experience-based career education in Perry-Parkdale Schools, Mid West. New York: Center for Policy Research, 1980.

Miles, M. B. et al. Final Report. Program on school architecture in education. New York: Center for Policy Research, 1978.

Parlett, M., & Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programs. Occasional paper No. 9. Evaluation Studies Review Annual, 1976, 1, 140-157.

Patton, M. Q. Qualitative evaluation methods. Beverly Hills, Calif.: Sage Publications, 1980.

Reichardt, C. S. On combining qualitative and quantitative methods in educational evaluation. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

Sadler, D. R. Intuitive data processing as a potential source of bias in naturalistic evaluations. Educational Evaluation and Policy Analysis 1981, 3(4), 25-31.

Schatzman, L., & Strauss, A. L. Field research: Strategies for a natural sociology. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Smith, L. M., & Geoffrey, W. The complexities of an urban classroom: An analysis toward a general theory of teaching. New York: Holt, Rinehart, and Winston, 1968.

Smith, L. M., & Pohland, P. A. Education, technology, and the rural highlands. In Four evaluation examples: Anthropological, economic, narrative, and portrayal. AERA Monograph Series on Curriculum Evaluation, No. 7. Chicago: Rand McNally, 1974.

Spradley, J. P. You owe yourself a drunk: An ethnography of urban nomads. Boston: Little, Brown and Co., 1970.

Stake, R. E. Progressive focusing. Paper delivered at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Stake, R. E., & Easley, J. A., Jr. Overview. Booklet 0. Case studies in science education. Urbana, Ill.: University of Illinois, Center for Instructional Research and Curriculum Evaluation, 1978.

Sutherland, A. Gypsies: The hidden Americans. London: Tavistock Publications, 1975.

Trankel, A. Reliability of evidence: Methods for analyzing and assessing witness statements. Stockholm: Beckmans, 1972.

Van Maanen, J. The fact of fiction in organizational ethnography. Administrative Science Quarterly, 1979, 24, 539-550.

Vidich, A., & Bensman, J. The validity of field data. In R. N. Adams & J. J. Preiss (Eds.), Human organization research: Field relations and techniques. Homewood, Ill.: Dorsey Press, 1960.

Wolcott, H. W. The man in the principal's office: An ethnography. New York: Holt, Rinehart, and Winston, 1973.

Wolcott, H. W. Teachers vs. technocrats: An educational innovation in anthropological perspective. Eugene, Ore.: Center for Educational Policy and Management, University of Oregon, 1977.

Whyte, W. F. Street corner society: The social structure of an Italian slum. Chicago: University of Chicago Press, 1955.

Yin, R. K. The case study as a serious research strategy. Knowledge: Creation, Diffusion, Utilization, 1981, 3, 97-114.

Yin, R. K., & Gwaltney, M. K. Organizations collaborating to improve educational practice. Washington, D. C.: Abt Associates, 1981.